

Published as: Deeb, R., Ooms, K., Vanopbroeke, V., De Maeyer, P. (2014). Evaluating the Efficiency of Typographic Design: Gender and Expertise Variation. *The Cartographic Journal*, vol. 51 (1), 75-86.

Evaluating the Efficiency of Typographic Design: Gender and Expertise Variation

Rasha Deeb, Kristien Ooms, Valerie Vanopbroeke, Philippe De Maeyer
Ghent University, Department of Geography
Krijgslaan 281, S8, B-9000, Ghent, Belgium
{Rasha.Deeb, Kristien.Ooms, Valerie.Vanopbroeke, Philippe.DeMaeyer}@UGent.be

ABSTRACT

Although the efficiency of label placement algorithms has been studied extensively, few studies considered the influence of the label designs on the efficiency of map readers. Labels are one of the most important elements on the map as they can provide more information than other symbols can. The design of the labels does have to stress the theme, shape, and functionality of the associated objects, which results in a more efficient interpretation of the map content by the user. How the label designs can enhance the map readers' efficiency (and thus the quality of the maps themselves) is the main objective of this study. A user study was conducted in which the participants were asked to locate a target label on a map. Different label designs were implemented across the trials. The participants' reactions times were registered to measure their efficiency and statistically analysed using a one-way ANOVA. Two different users' characteristics were considered: gender and expertise. Related to the size, shape, orientation, and texture of the labels, a number of significant differences ($P < 0.05$) and trends were located. Differences in efficiency between males and females, on the one hand, and between novices and experts, on the other hand, were also described statistically. Consequently, recommendations can be formulated regarding the design of labels in order to obtain more efficient maps, keeping in mind the map users' characteristics.

Keywords: Visual variable, label efficiency, label design

INTRODUCTION

Cartography comprises a systematic symbolization of the features in the environment, nevertheless the addition of labels is crucial to finalize the cartographic product. Gerber (1981) stressed the importance of labels by identifying three subsequent levels for a successful cartographic interpretation ranging from the simplest to the most complex: the

perception-recipe level (pictorial), the label level (pictorial-verbal), and the 'other knowledge about' level (verbal). In the midterm level, the label level, the map reader gains knowledge about the name of the object he tries to interpret. Consequently, labeling objects on maps contributes significantly to the users' interpretation process. This facilitates the link between the pictorial and verbal level, which is critical to interpret the map correctly. Imhof (1975) demonstrated that map lettering is a linguistic, practical, technical, and esthetic issue. Regarding the use of these labels on maps, two main issues can be identified. The first issue concerns the placement of the labels, both on the algorithmic level known as automated placement, and on the cartographic level (quality of the placement) described by Imhof (1975) as good or poor placement. The second issue concerns the actual design of the labels: how should they be presented in order to create maps with a higher quality and which can thus be interpreted more efficiently by the map users.

Label Design

The design, or representation, of map labels is strongly connected to the function and type of associated feature divided into position, linear, and areal designation (Imhof, 1975). Kraak and Ormeling (2010) differentiated between a primary label function of providing the geospatial addresses and a secondary function of indicating the nature of the represented object. For this they set rules to indicate hierarchal and nominal differences by changing the variables in which the label is presented. They proposed using the variation of boldness, size, spacing, colour value, and/or case style to represent hierarchal differences. They also proposed the use of color, shape, and straight script vs. italic script to represent nominal differences. In addition to that, Krygier and Wood (2011) also set some functional rules for using type as a map symbol such as using different typeface to indicated different qualitative data and using size and type weight to present order. Deeb *et al* (2012) linked the use of visual variable to the perceptual characteristic of label design (associativity, selectivity, order, and quantity). The density of names on a map depends on the content and purpose of the map. In order to facilitate the interpretation of the visual information on maps, Imhof (1975) set rules to avoid type accumulation, type overlapping, wrong orientation, and type spreading, then he demonstrated the drawbacks of bad designs. Peterson (2009) discussed the use of uppercase and lowercase and suggested that lowercase letters are read easily because of the different height of various letters resulting in a specific shape for each letter; meanwhile the uppercase letters have the same height and global shape which make them harder to read. Bartz (1970a) pointed out that applying typographic findings reported in non-cartographic literature is acceptable because of the similar functional task that cartographic labels deliver. In order to set empirical rules, Bartz (1970b) investigated typographic variables of both shape and size on paper maps to evaluate their influence on search time for both individual label and the complete searching task. Different design criteria were also investigated by Phillips (1981) as he tested the influence of character design regarding the word's initial letter, the word length, and the word shape. By using eye tracking method, he concluded that these elements affect the fixation time for the target names. The association with the typographic design function and the readability of the text on the map was discussed by Fairbairn (1993).

Wood (2000) described the functions of label design and set rules for principle positioning of lettering according to their own function, illustrating the optimal typographic

design according to the shape of depicted features. To investigate the influence of the labels' shape on screen map readability, Feldmann and Kreiter (2006) conducted a controlled test on three sans serif fonts (Arial, Univers, and Frutiger), but they found no significant difference between the readability of the three of them. Before that, Arditi and Cho (2005) investigated the legibility of serif and sans serif fonts on a plain text and concluded that serif fonts is slightly more legible than sans serif fonts. The readability of labels is expected to be affected by the typographic design (van den Worm, 2001: 87-107). He proposed that bold typography could improve readability but this may clash with desirable perception characteristics. Ever since, no empirical proofs supported or disapproved this proposal.

Map Usability Testing

The development of the cartographic products was combined with a growing interest of assessing their usability; including map effectiveness, map efficiency, and map satisfaction (Faulkner, 2000). The usability of interactive mapping tools was tested by Andrienko *et al.* (2002) where they tested the tool learnability, memorability, and user satisfaction. Their study concluded that users were able to adopt the new ideas for map interactivity and manipulability. Nivala *et al.* (2008) tested the functionality and the features of web mapping sites, they found out some usability problems related to search operation, user interface, map visualization, and map tools. To avoid such problems they suggested some design guidelines for web mapping design. Effectiveness of map design methods were tested by MacEachren (1982) when he examined the effectiveness of both choropleth maps and isopleth maps for direct data acquisition and pattern memory, exploring the role of complexity in their design. He concluded that the isopleth maps are more effective in the term of memorizing general pattern. The design variable were also tested by Garlandini and Fabrikant (2009), they reached empirical results suggesting that size is the most efficient and effective visual variable to detect changes under flicker conditions and orientation proved to be the least efficient and effective visual variable among size, hue, value, and orientation. Users' preference (satisfaction) of label design associated with point and areal data was tested by Deeb *et al.* (2012). They applied Bertin's (1970) visual variables on labels and investigated the users' preference towards different label designs and concluded some aesthetics design criteria. In addition to that, they indicated significant differences between user groups which were tested: novices vs. experts and females vs. males.

Bearing in mind that the aesthetics of the label design, tested through users' preference, does not always reflect an optimal design. The main goal of a cartographic product is communication: to get a message across as efficient as possible. With efficiency, the facility with which a user can interpret the content is meant. Therefore, it is crucial to complete the aesthetics measurements with information about the efficiency of certain label designs towards different map users. In order to measure the efficiency with which the map content is interpreted, reaction time measurements are often used. The latter issue is the main concern of the user study described in this paper.

Quantifying users' performance can be done either by the numbers of achieved tasks or the time user takes to complete specific task (Nielsen, 1993). Measuring reaction times (RT) is often used to answer questions regarding the users' cognitive process. A fundamental method to measure performance is measuring objective time to complete a task (Rubin and

Chisnell, 2008). Measuring reaction times does not only reflect the mental processing speed, but is also linked to how fast the information is retrieved. The users reaction times consist out of the total time needed to execute an assignment including the cognitive processes linked to it: analysing, storing, recoding, and subsequently using the raw data. In this paper the users' reaction time to find a certain label on a map, linked with a number of label designs, is used as a measurement of their efficiency. The time consumed to find the target labels reflects the effect that different label designs have on the users' efficiency to read the map contents. Typically users pass by four stages to locate their targets. First stage is identifying the target, then memorizing the name, third is to adjust the typography of the target to the typography used in the map and finally searching to find the target. The structure of the study ensures that all users can pass through these stages without any interfering elements that could cause a bias to the results. Due to the fact that user characteristics should be captured and taken into account during the design (Haklay and Nivala, 2010), the study has both a between- and within-user design as the efficiency of different label designs were tested, considering different user characteristics (gender and expertise). The design of the user study is described in detail in the next sections.

STUDY DESIGN

Participants

Two types of user characteristics are studied in the experiment, gender and expertise, which is reflected in the selected participants. The novice group included 25 participants with an average age of 16.4 years. This group consists of pupils at the secondary school level and, consequently, with no previous training in cartography. The expert group consisted of 25 participants who work with maps on a daily basis and have at least a master's degree in geography. The average age within the expert group was 29.5 years. Of the 50 participants, 25 were female and 25 male. The average age of both females and males was 23.

Task and Stimuli

A series of forty maps was presented to each participant in an online questionnaire. Both areal and point data were integrated in the map design, represented by thematic and topographic maps. Two examples of such maps used in the experiment are illustrated in Figure 1. For each map the participants had to locate a target label in the map image. They were instructed to click on the label in the map image when it was found, which resulted in a reaction time measurement (time interval, in milliseconds, between the display of the map and the mouse click). After the mouse click, a new map and new target label was displayed. These instructions were formulated on the first screen of the test. After the completion of the forty maps, the participant had to fill in a post study questionnaire. Gender, age, level of expertise, and other personal characteristics were registered in this questionnaire.

The target label was displayed above each map and was depicted in a neutral font (OCR A Extended) that was not used in the map image itself. The length of the target labels was kept constant: 6 or 7 letters. The names were carefully chosen regarding their general shape, taking into account, among others, their cap height, x-height, and ascenders. The location of the target label within the maps was chosen carefully, based on eye movement

analysis from Ooms *et al.* (2012). Visual search strategies can also vary between users as Lleras and Mühlénen (2003) clarified. They discovered that different search approaches were followed. The search strategies were divided into either a systematic approach, as users geometrically scanned the map (top to bottom, bottom to top, right to left, and left to right), or an intuitive approach as the users randomly searched the map image. In order to level out these search approaches, target labels were distributed equally on all direction and covered the four corners (see Figure 2). Labels design on the map varied in size, shape, orientation, and texture, which were applied individually or combined. An extensive explanation of these visual variables can be found in the work of Bertin (1970) and their application on map labels was described by Deeb *et al.* (2012). Fictive labels were placed on the map to assure that participants would not recognize the location. This could bias the results due to previous knowledge: a user can retrieve the position of a label much faster if he is familiar with the region.

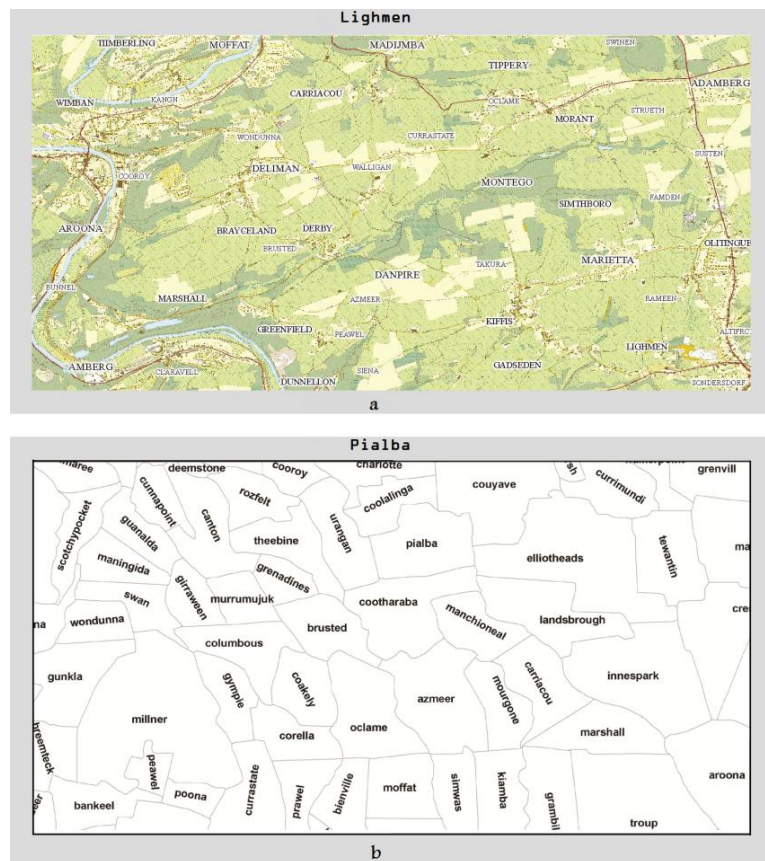


Figure 1: An example of two test screens, (a) point data labelling, (b) areal data labelling

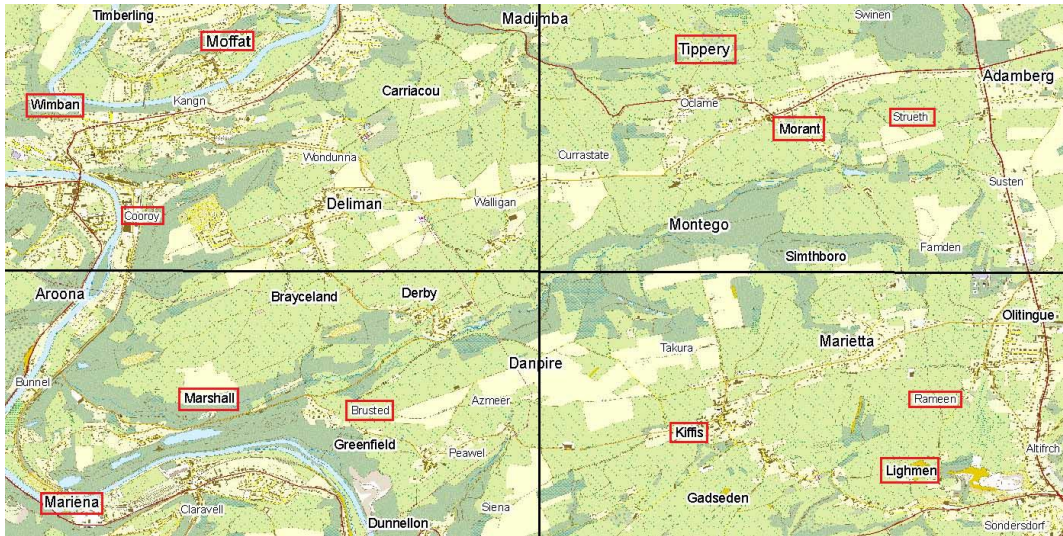


Figure 2: An example of label target distribution on the topographic map

To avoid biases in the measurement due to resolution and size differences, all participants completed their test on a flat screen with a 1280x1024 resolution. The latency of mouse registration time and the software registration time were taken into account. Furthermore, the test was run on hardware with similar properties and presented in the same browser (Google Chrome). Each map had a similar size (about 350 KB) to ensure equal loading times of it. Each participant followed the same order of maps in a sequence that lasted 10- 15 minutes. The same label was never asked more than twice and the studied variables were presented as such that maps related to the same variable were not presented directly after each other.

RESULTS AND DISCUSSION

Data

All data were collected through an online questionnaire and stored in a database. However, not all participants completed the questionnaire until the end, and consequently, no personal characteristics could be gathered. In order to assure the consistency of the dataset, all incomplete records were removed from the database. Furthermore, out of 2000 measurements 54 outliers were detected in the reaction time measurements. All these outliers were larger than the mean value plus two times the standard deviation ($M+2 SD$; indicating the limit of the 95% confidence interval). These outliers could be explained by distractions of the participant during the experiment. If the participant is distracted at a certain point, the reaction times would show a steep increase. However, these values do not contribute to the goal of our research and they are a cause of distortion in the complete dataset. As a consequence it was decided to remove these outliers from the dataset. No outliers less than $[M-2 SD]$ were found which support the theory of participant distractions related to the other outliers.

Size of Labels

Using Arial font, four label sizes were integrated in the test: 8, 10, 12, and 14 points. In addition to that, a bold typography was applied to the previous label sizes, which even enlarges the characters. The users' reaction time measurements were collected and a one-way

ANOVA test was applied to both groups of gender and expertise. In order to get an overview of the mean values and distributions of the registered reaction time, Figure 3 illustrates the box plots of each category and group. From this figure it can be derived that the reaction times do vary between the user groups, however no systematic trend can be observed for the four different sizes. The shape of the box plots represents the range of reaction time observations and it can be noticed that the range of the reaction time measurements related to a normal font was larger than these related to a bold font. This remains true for both gender and expertise sub-groups.

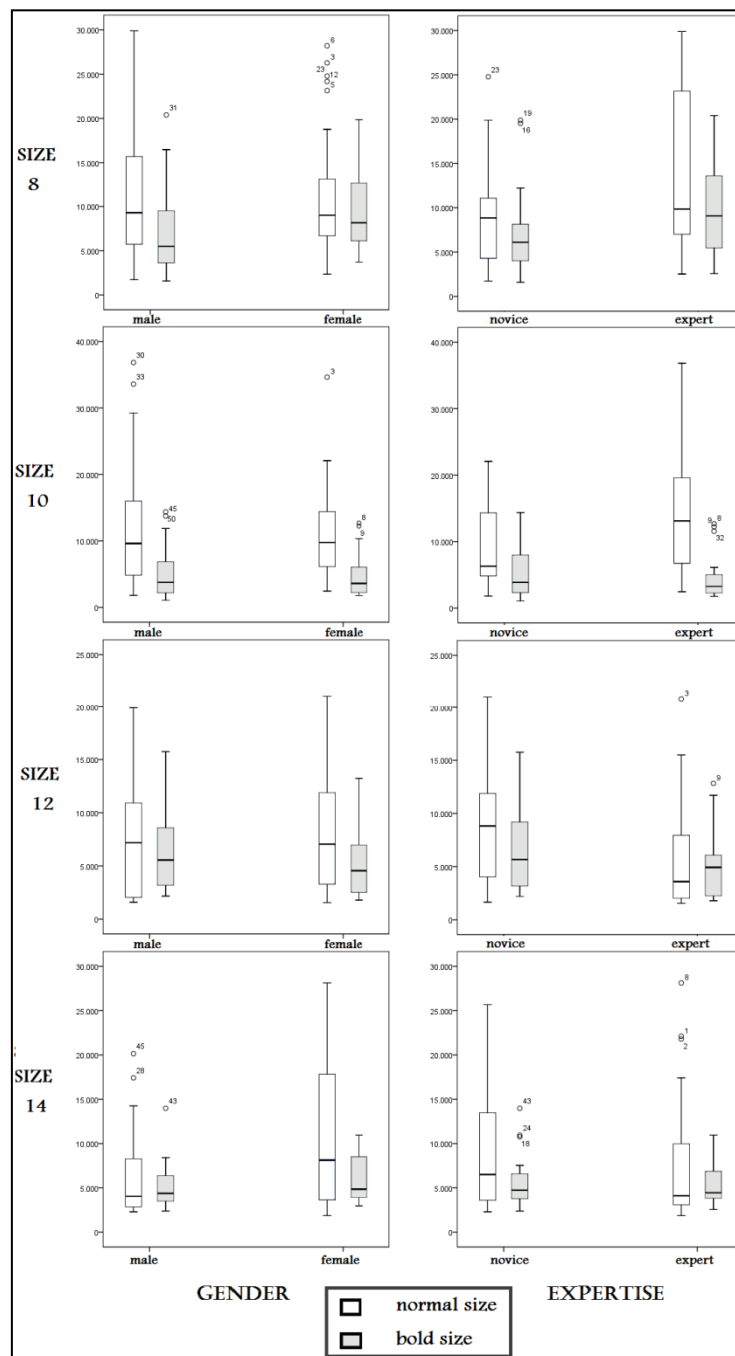


Figure 3: Box plots of the reaction time measurements (in ms) related to the labels' size, boldness, and user groups

Table 1 lists the mean reaction times (M) and associated standard deviations (SD) for males and females, related to the different label sizes. This table shows a trend in the reaction time measurements related to the bold and normal design of labels: both user groups were numerically faster in finding the labels with a bold design in comparison to the normal label design. A one-way ANOVA test on reaction time measurements of normal size vs. bold size for gender (males vs. females) demonstrates this difference is highly significant ($F=13.600$, $P=0.002$). Table 1 also indicates decreasing reaction times when larger labels sizes were used for the males. However, the labels with size 12 points in bold caused a deviation in this trend. In the females user group, more deviations from this trend were observed. A one-way ANOVA shows that the males' and females' reaction time measurements were very similar: only the reaction times related to locating labels with a size of 14 points (normal) were significantly shorter for the males ($M=6354$ ms) than females ($M=10566$ ms). As only one value near to significant difference demonstrate the variation between the males and females for the size 8 bold. The most efficient label size seemed to correspond to bold 14 points for the males ($M=4873$ ms) and bold 10 points for the females ($M=4794$ ms).

Table 1: Mean reaction times (in ms) for males and females, related to label sizes

Size (points)		Males		Females		ANOVA	
		M	SD	M	SD	F	P
8	normal	11464	8133	11564	7901	0.002	0.965
	bold	6976	4892	9670	4984	3.719	<u>0.060</u>
10	normal	12320	9931	11048	10294	0.260	0.613
	bold	5306	4033	4794	3320	0.221	0.640
12	normal	7099	5408	8094	5976	0.381	0.540
	bold	6201	3722	5626	3679	0.303	0.585
14	normal	6354	5036	10599	8533	4.161	0.047
	bold	4873	2460	6119	2808	2.732	0.105

The results, listed in Table 3, present an overview of the mean (M) reaction time measurements and standard deviations (SD) for the novices and experts. The last two columns show the results from the statistical comparison between the two groups of users, using a one-way ANOVA. This table shows a similar trend as with the males and females: all users (novices vs. experts) could find bold labels faster than normal labels. Furthermore, this difference can be considered highly significant ($F=10.670$, $P=0.006$). This means that the bold design of labels was more efficient than the normal design towards these different user groups. Furthermore, the results from the ANOVA test revealed a similar level of efficiency between the two groups of users. However, two significant differences were measured between both user groups. Novices could locate the labels with a size of 8 points (normal) much faster (and thus efficiently) than the experts. In contrary, the experts could locate the labels with a size of 12 points (normal) more significantly efficient than the novices. The novices considered the bold labels with a size of 14 points as the most efficient design ($M=5532$ ms), however only one test showed a near to significant different value with a size of 10 points was noted ($M=5594$ ms). The experts could locate the labels with the latter design (bold, 10 points) most efficiently ($M=4516$ ms).

Table 2: Mean reaction times (in ms) for novices and experts, related to label sizes

	Size	Novices		Experts		ANOVA	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
8	normal	9155	6031	13874	8978	4.758	0.034
	bold	7156	4684	9490	5271	2.737	0.105
10	normal	9293	5883	13954	10294	0.613	<u>0.059</u>
	bold	5594	4015	4516	3305	1.004	0.322
12	normal	9201	5697	5993	5260	4.279	0.044
	bold	6581	4020	5246	3236	1.672	0.202
14	normal	9164	7440	8084	7399	0.249	0.620
	bold	5532	2908	5437	2508	0.015	0.903

The experiment indicate that all users groups have a significant better performance when using bold size as their reaction time measurements were significantly less than those of normal size measurements. Matching reaction time measurements to the four tested font size did not show any systematic trend to predict the efficiency according the increase or decrease of font size. Nevertheless, it showed typical cases where both groups agreed with font size (both point size and boldness) and other cases of disagreement. Variations in size correspond to variations in performance according to both the gender group and the expertise group. The four tested size of Arial font (8, 10, 12, and 14) showed different trends but what is interesting about these trends is that all comparisons agreed on more efficient performance for bold design in the four tested sizes. In addition to that, the bold design did not witness any significant difference for the different point sizes and the different sub-groups. This results agrees with van den Worm (2001) who suggested that bold typographic design is more readable than the normal typographic design. In addition to that, it also contributes to what Deeb *et al.* (2012) presented in their research of users' preference towards label design as they found out that bold size design is more preferred than the normal size design.

Shape of Labels

The shape of label is mainly determined by its font: its typeface at a certain size. Theoretically, the world of shapes is infinite, which thus hold also true for fonts (Bertin 1970). In this infinitive amount of fonts, different families can be identified, from which 'serif fonts' and 'sans serif fonts' are two of the most important ones. The efficiency of these two font families was tested using two of its most used (and known) typefaces, namely Arial (sans serif) and Times New Roman (TNR, serif). Both fonts were presented at a size of 10 points (normal) and on a blank background to avoid any disturbance of the surrounding elements and influence of colours on the user's cognitive load.

Table 3 lists the mean (*M*) reaction times of the males and females and Table 4 lists the mean reaction times of the novices and experts. The last two columns of both tables contain the results of the ANOVA tests between the sub-groups. These tests indicate that the efficiency with which the labels were located is similar in all user groups. Only one value near to significant difference was detected between novices and experts for Arial font.

Table 3: Mean reaction times (in ms) for males and females, related to the labels' shape

Font	Males		Females		ANOVA	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
TNR	7021	5047	6530	3899	0.140	0.710
Arial	12320	9931	11048	7408	0.260	0.613

Table 4: Mean reaction times (in ms) for novices and experts, related to the labels' shape

Font	Novices		Experts		ANOVA	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
TNR	7086	4292	6486	4758	0.210	0.649
Arial	13954	5883	9293	10294	3.744	<u>0.059</u>

Analysing the mean reaction times based using one-way ANOVA test showed no influence on the use of a serif or sans serif fonts on both gender ($F=0.546$, $P=0.537$) and expertise ($F=4.238$, $P=0.176$). Both the between- and within-user reaction time study agreed on the efficiency of utilising serif and sans serif fonts. Further designs of serif and sans serif fonts on coloured backgrounds were undertaken to demonstrate the influence of case style variation on users' efficiency of both Times New Roman and Arial fonts: all lowercase, only first uppercase, all uppercase. The results of these six different designs are depicted in Table 5 (males vs. females) and Table 6 (novices vs. experts).

In both tables, the reaction time measurements seemed to be rather similar between the two user groups. Only one significant difference could be detected between males and females: females could locate the labels in the 'First letter uppercase, Arial' design significantly faster than the males ($F=6.155$, $P=0.017$). No general trend could be observed between the use of the Arial and Times New Roman font regarding the efficiency of the users. However, one remarkable trend was noticed in the 'all lower case' design. The largest reaction time measurements for all user groups were registered with the Times New Roman font, whereas the smallest reaction time measurements were linked with the Arial font. This would mean that this design is the least efficient when the Times New Roman font is used, and it is at the same time the most efficient when the Arial font is applied. The expertise group showed two values near to significant when ANOVA test was applied to test the significant variation between the user sub-groups. Both values were recorded in the category of all letters lower case. For TNR font novices ($M=10491$ ms) performed the task faster than experts ($M=14342$ ms) and this was near to significant difference ($F=3.695$, $P=0.061$). For Arial font experts ($M=5844$ ms) performed the task faster than novices ($M=6976$ ms) and this also was near to significant ($F=3.060$, $P=0.087$).

Table 5: Mean reaction times (in ms) for males and females, related to categories of labels using case styling

Categories	Font	Males		Females		ANOVA	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
all lowercase	TNR	13001	6469	11832	8094	0.138	0.575
	Arial	6587	2249	6233	2365	0.282	0.598
first uppercase	TNR	6875	3203	6738	3742	0.018	0.895
	Arial	9548	5783	6384	2683	6.155	0.017
all uppercase	TNR	8439	4400	7968	4120	0.141	0.709
	Arial	8482	4152	10092	7826	0.785	0.380

Table 6: Mean reaction times (in ms) for novices and experts, related to categories of labels using case styling

Categories	Font	Novices		Experts		ANOVA	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
all lowercase	TNR	10491	6448	14342	7666	3.695	<u>0.061</u>
	Arial	6976	2298	5844	2184	3.060	<u>0.087</u>
first uppercase	TNR	6857	3668	6752	3245	0.010	0.920
	Arial	7024	3709	8908	5500	2.016	0.162

Categories	Font	Novices		Experts		ANOVA	
all uppercase	TNR	7231	4055	9021	4265	2.101	0.154
	Arial	10043	7608	8528	4539	0.695	0.409

To obtain general conclusion about the efficiency of user groups towards the two fonts for the described design, a one-way ANOVA test was used to describe statistically the efficiency of TNR and Arial. The gender sub-groups showed no significant difference in their reaction time measurements of TNR and Arial in the three described case-style ($F=0.952$, $P=0.352$) and so did the expertise sub-groups ($F=0.819$, $P=0.352$).

Notwithstanding its general character, the results reported here would seem to indicate that the efficiency of the studied font family (serif, sans serif) did not affect the user reaction time measurements. However, the case styling of the labels in combination with a certain font family did show a significant influence on the user's efficiency between males and females. It is critical to establish rules for the shape design efficiency spatially when the design is used for different function. this topic was researched thoroughly preference wise by Deeb *et al.* (2012) when they tackled the case style, they concluded that 70-80 % of users preferred Arial font over Times New Roman. The participants' efficient performance on Arial contradicts Bartz (1970a), who suggested applying typographic findings that have been reported in the non-cartographic literature. Because this result disagree with Arditi and Cho (2005) who concluded from their experiment on plain text that serif fonts are more legible than sans serif fonts. The result considering the efficiency of case style could be explained as Peterson (2009) demonstrated; it is easier to read the lowercase letters because they have different height and thus different shapes, unlike the uppercase letters which all have the same height and makes the process of distinguishing letters longer.

Orientation of Labels

The orientation of the labels on a map can be considered at two levels: orientation of the whole label and orientation of the characters within the label. Studying the first level, labels were placed according to three approaches. First approach (Horizontal) considers placing all labels horizontally. The second approach (Tilted) implies placing the label under an angle, based on the shape of the object. In this case, label will stress the general shape of the object, as the labels are placed along its main diagonal. The third approach (Mixed) gives a higher priority to horizontally placed labels. Only when this is not possible, the labels will be tilted. These three types of label orientations were integrated in the test. The results of the mean reaction time measurements (M) and the statistical comparison between the different user groups are presented in and Table 7 (males vs. females) and Table 8 (novices vs. experts).

In the gender comparison, one significant difference was detected, related to the horizontal orientation of the labels ($F=5.685$, $P=0.021$). Females ($M=6768$ ms) could locate these names faster (and thus more efficiently) than the males ($M=9785$ ms). In the case of the novices vs. experts comparison, also one significant difference was detected but this time for the tilted approach ($F=6.314$, $P=0.015$). The novice participants ($M=8428$ ms) seemed to be able to located the tilted labels more efficiently than the expert participants ($M=14868$ ms).

No systematic trend could be observed in the data. Females and novices were numerically faster in locating the horizontally placed labels than with tilted and mixed orientation. However, the males and experts showed little difference in the mean reaction

times between the horizontal and tilted orientation, and a larger difference with the mixed approach. However, none of the mean reaction times pointed to the mixed design as the most optimal label orientation (in terms of efficiency).

Within-users analysis and under the experiment condition, using one-way ANOVA did not show any significant difference of the reaction time measurements of the orientation design for both gender group ($F=3.063$, $P=0.188$) and expertise group ($F=1.678$, $P=0.324$). However, it would be interesting to extend this work to cover different functions of the orientation placement. This result agrees with Garlandini and Fabrikant (2009) as they found that the visual variable orientation proved to be the least efficient and effective visual variable. The three orientations did not cause any significant difference between the users sub-groups of expertise and gender. The trend of user performance in reaction time tasks agrees with the trend that Deeb *et al.* (2012) introduced, when they measured users preference towards the orientation of label design. Horizontal orientation was the most preferred orientation over the tiled and mixed orientation

Table 7: Mean reaction times (in ms) for males and females, related to the labels' orientation

Orientation	Males		Females		ANOVA	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
Horizontal	9785	5188	6768	3213	5.685	0.021
Tilted	9626	7924	13460	10534	2.059	0.158
Mixed	13090	10080	13121	10546	0.432	0.514

Table 8: Mean reaction times (in ms) for novices and experts s, related to the labels' orientation

Orientation	Novices		Experts		ANOVA	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
Horizontal	7895	4208	8705	4916	0.367	0.548
Tilted	8428	5736	14868	11402	6.314	0.015
Mixed	12511	9899	13724	10686	0.936	0.338

When considering the second level of label orientation, the individual characters are tilted. This is achieved by using an italic (Arial) font. The results of the normal font (straight), are listed in Table 1 and Table 2, and with those presented in italic are listed in Table 9 and Table 10. In order to obtain a structured overview, the box plots of these measurements are depicted in Figure 3. From this figure it can be derived that the mean reaction times are systematically higher when a label in a straight font had to be located. The mean reaction time measurements for the gender sub-groups indicated that labels in the italic design could be located significantly faster than in the straight design (ANOVA, $F=21.549$, $P=0.000$), and thus more efficiently. Similar results were found for the expertise sub-groups (ANOVA, $F=15.841$, $P=0.001$). This could be due to the attraction that italic causes as it is not the typical design of labels and usually used to make distinction of thematical purposes.

Using Arial font, four italic sizes were tested using ANOVA, which indicated no significant difference in both gender and expertise groups (see Table 9 and Table 10) as ($P > 0.05$). A trend was noticed; for the smallest size (8 and 10) males were faster than females meanwhile females were faster for the larger size (12 and 14). Novices were faster for the smallest size 8 and experts were more efficient than novices for the larger sizes (10, 12, and 14). Only one value near to significant was recorded and it was for size 10 italic in the expertise group

($F=2.904$, $P=0.095$). As italic showed significantly better users performance, it could be used as a powerful element to indicate importance of some map elements.

Table 9: Mean reaction times (in ms) for males and females, related to the use of italic

Italic	Males		Females		ANOVA	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
Size 8	7028	3109	7215	4230	0.029	0.866
Size 10	4683	3209	5141	2670	0.288	0.594
Size 12	5203	3387	4728	2454	0.319	0.575
Size 14	6261	2856	5581	1726	1.016	0.319

Table 10: Mean reaction times (in ms) for novices and experts, related to the use of italic

Italic	Novices		Experts		ANOVA	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
Size 8	6155	2944	7940	4109	2.765	0.103
Size 10	5649	3686	4234	1831	2.904	<u>0.095</u>
Size 12	5337	3460	4599	2321	0.775	0.383
Size 14	5930	2066	5883	2624	0.005	0.946

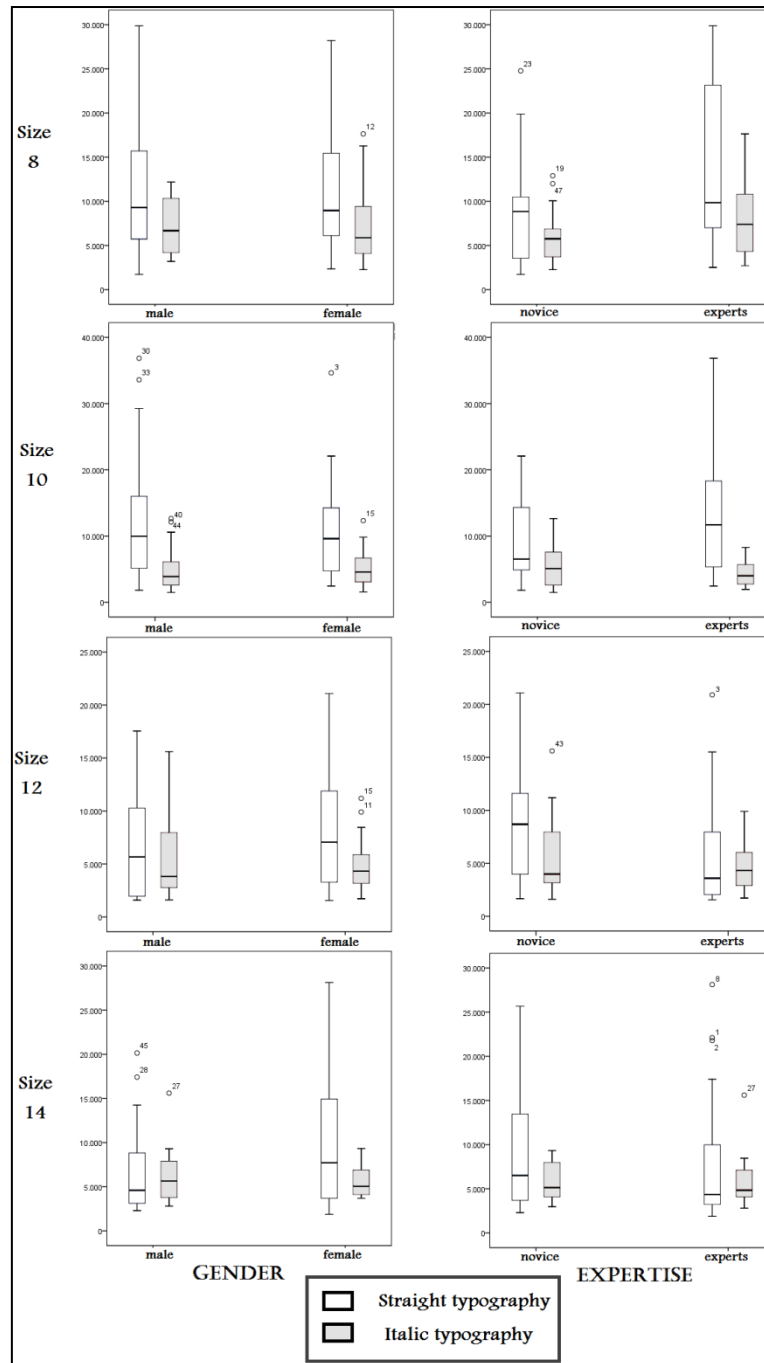


Figure 3: Box plots of the reaction time measurements (in ms) related to the labels' orientation, straight vs. italic and user groups

Texture of Labels

The design of the labels' texture has a wide array of choices as they are based on a combination of visual variables (Deeb *et al.*, 2012). Using textures, different categories of labels can be visualised by combining different label designs (one design for each category). These categories might correspond to different hierarchic levels in the labels (e.g. city vs. village) or functional classes (rivers vs. roads). Hierarchic levels can be presented by variations in the labels' size. In this study, three consecutive sizes were used (12, 10, and 8 points) to present three levels of hierarchy. Size of 12 points and 10 points were used to present two levels of hierarchy. Both levels of hierarchy were presented in a Times New

Roman (TNR) and an Arial font. In Table 11, the mean (*M*) reaction time measurements for the males and females users are listed, together with their mutual statistical comparison using ANOVA. Table 12 displays these results for the novice and expert users.

The lowest reaction times were linked with three levels of hierarchy in Times New Roman. However, these measurements did not deviate much from these linked with the Arial font. Larger deviations were found between the designs with respectively two and three levels of hierarchy. Longer search times were registered when only a '2 Levels' hierarchy was depicted. One exception on this was the mean reaction time measurements of the novices for locating labels when a '2 Levels' hierarchy was depicted in an Arial font. This latter combination resulted in a mean search time which was rather similar to the one of the '3 Levels, TNR' label design. This was not the case within all other user groups. One significant difference was located in the group of expertise for the '2 Levels, Arial' ($F=6.982$, $P=0.011$) when novices performed significantly faster ($M=7768$ ms) than experts ($M=13029$ ms).

A one-way ANOVA test showed a statistical difference between both categories (2 Levels vs. 3 Levels) when males and females were considered ($F=10.196$, $P=0.019$). However, there was not any statistical difference between both categories when expertise groups were considered ($F=2.628$, $P=0.156$). The gender group indicated that the labels in the 3-levels design could be located more significantly efficient than in 2 level design (see Table 11). Table 12 shows that there is no trend to the expertise sub-groups reaction time measurements which indicates the need for further research in this domain. No systemic trend was detected for the influence of complexity on users' efficiency. But for the gender sub-groups. There was a trend addressing the role of functional complexity; the higher the complexity (3 level), the more efficiency the design is. This result disagree with MacEachern (1982) conclusion on the influence of complexity effectiveness wise on data direct acquisition, considering isopleth and choropleth symbolization methods. Because he found out that no significant difference of users performance towards the two symbolization methods.

Table 11: Mean reaction times (in ms) for males and females, related to categories of labels using font sizes

Categories	Font	Males		Females		ANOVA	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
3 Levels	TNR	8797	4509	8040	4620	0.337	0.564
	Arial	9661	5113	9827	6031	0.011	0.918
2 Levels	TNR	12842	5589	11613	5651	0.573	0.453
	Arial	10629	8037	10282	6869	0.027	0.871

Table 12: Mean reaction times (in ms) for novices and experts, related to categories of labels using font sizes

Categories	Font	Novices		Experts		ANOVA	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>
3 Levels	TNR	7818	4663	9010	4417	0.844	0.363
	Arial	9414	5764	10091	5404	0.180	0.674
2 Levels	TNR	11078	6070	13236	5022	1.813	0.185
	Arial	7768	4190	13029	8844	6.982	0.011

CONCLUSIONS AND FUTURE PERSPECTIVES

The efficiency of map labels involves much more than creating legible labels. In addition to the optimal label placement, the functional design of the labels should be provided (Fairbairn, 1993). Four types of label designs were included in a user test, which are used to indicate thematical, hierarchical, and shape of the cartographic information (Deeb *et al.*, 2012).

Comparing the reaction time measurements for both groups, bold *size* does not carry any significant difference for both gender and expertise group. However, a trend was noticed here: all users could locate the labels more efficiently when they were presented in bold as opposed to not-bold (normal). The efficiency of the users across the different groups seemed to be very similar, from which it can be concluded that the design of the labels should not be made for each user group specifically. However, generally speaking, all users seem to be able to locate names more efficiently when the size of the labels (in points) increases. This can be explained by the fact that the labels will become better readable with a larger size. This is compatible with the general rule that van den Worm sat (2001) saying ‘text to be rasterised should not be smaller than 10 point’. Thus, it is not recommended to use the size 8 point for the mentioned reasons.

What is interesting about bold size is that it proved its efficiency over the normal size for both groups. Using bold size shows significantly faster reaction time than the normal size for gender group and so do for the expertise group. In addition to that, the findings recommend the use of bold as it was significantly more efficient than the normal size. This also corresponds to what van den Worm suggested (2001) of using bold font. Furthermore, it also matches the findings of Deeb *et al.* (2012) as they concluded that the use of bold font is the preferred use over the use of normal font. It should be noted that changing the font changes the type size (van den Worm, 2001:87-107), therefore these results are limited to the use of Arial font and could be extended to cover the sans serif font family based on Feldmann and Kreiter (2006) exploratory conclusions.

Variation in label *shapes* does not indicate any significant difference of using serif and sans serif font. In addition to that, there is no significant statistical difference between males vs. females and novices vs. experts except for the gender study of Arial first letter upper case. It is noteworthy that the results are more consistent when only one case-style is used. Although these results do not fully solve the arguments about the use of label shapes, it leads to a coherent description of serif and sans serif design when users’ characteristics are paired in mind.

Orientation variable shows a significant difference located in the gender group for the horizontal orientation (females were faster). In addition to that, the reaction time indicates a significant difference in the expertise group for straight orientation (novices were faster). Whereas no significant difference was detected for the mixed orientation in both gender and expertise study. The orientation of the label forming letters is expressed by italic. The use of the four sizes with italic do not show any significant difference in both gender and expertise groups. But it proved significantly its efficiency over the use of normal size and this could be explained by the attraction that the italic cause to the map users.

Hierarchical and thematical functions were used as a guide to design *texture*. Only one significant difference was deducted for hierarchy designs which use two level of hierarchy and Arial shape. When comparing the two designed hierarchy systems using ANOVA test, it

shows that the hierarchy group composed of three levels is significantly more efficient than the hierarchy of two levels for the gender groups, meanwhile there is not any significant difference of both hierarchy when studying the performance of expertise group. When comparing these results with what Deeb *et al.* (2012) presented, it worth mentioning that users always prefer the use of two levels of hierarchy. However, this (exploratory) study may offer some insight into adequate hierarchy presentation in term of users' efficiency. Further analyses need to be undertaken to indicate the optimal hierarchal design for different labelling purposes.

These conclusions can be employed when designing web sites with cartographic product or web sites for cartographic use. If the web site is open to all visitors, designers shall avoid using the variables in which significant differences between users were located. They can use variables where results were more consistent such as bold and italic. For more efficient design, designer can customise a log in to the user interface to determine users' characteristics such as gender and level of experience. Consequently, they can use the most efficient variable for each function and group as described earlier. This can also be employed when designers know the audience characteristics; such as making maps and atlases for research use where they can use the results of expert reaction time which indicated the most efficient variable. Such products can also help students at schools by making maps for gender differences.

Future work of tracing the performance of map audience according to the label design variation is a crucial for both the map users and the efficiency of labels. Different age group response to label design will be studied as well as different map use which seems crucial to accomplish this work. In addition to that, labelling linear features and the properties of their design will be tackled. Furthermore, the efficiency of other alphabetical systems such as Arabic and Chinese will be studied and thus, a comparison with the Latin design finding could be made. Moreover, the visual variable of colour and value will be involved in label design with a suitable map use and suitable labels' function.

REFERENCES

- Andrienko, N., Andrienko, G., Voss, H., Bernardo, F. Hipolito, J., and Kretchmer, U. (2002) 'Testing the Usability of Interactive Maps in Common GIS', **Cartography and Geographic Information Science**, 29(4), pp.325-342.
- Arditi, A. & Cho, J. (2005). Serif and font legibility. **Vision Research**, 45, 2926-2933.
- Bertin, J. (1970) **Semiology of Graphics: Diagram, Networks, Maps**, Translated by William J. Berg, Esri press, California.
- Bratz, B.S. (1970a) 'An Analysis of the Typographic Legibility Literature, Assessment of Its Applicability to Cartography', **The Cartographic Journal**, 7(1), pp. 6-17.
- Bratz, B.S. (1970b) 'Experimental Use of The Search Task in The Analysis of Type Legibility in Cartography', **The Cartographic Journal**, 7(2), pp. 103-113.
- Deeb, R., Ooms, K. De Maeyer, P. (2012) 'Typography in the Eyes of Bertin, Gender and Expertise Variation', **The Cartographic Journal**, (*accepted March 2012*).
- Fairbairn, D.J. (1993) 'On The Nature of Cartographic Text', **The Cartographic Journal**, 30(2), pp. 104-111.
- Faulkner, X (2000) **Usability Engineering**. Palgrave, New York.

- Feldmann, H-U. and Kreiter, N. (2006) 'Neuaufbau der schweizerischen, Landeskarte: Inhalt und Kartengrafik', **Kartographische Nachrichten**, 3, pp. 115–121.
- Garlandini, S., and Fabrikant, S.I. (2009) 'Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization'. In: **Lecture Notes in Computer Science, Spatial Information Theory**, ed. by Hornsby, K.S. et al., 5756, pp. 195-211, Springer-Verlag, Berlin Heidelberg.
- Gerber, R.V. (1981) 'Competence and Performance in Cartographic Language', **The Cartographic Journal**. 18 (2), pp. 104-111.
- Haklay, M. M. and Nivala, A. M (2010) 'User-centred Design', In: *Interacting with Geospatial Technologies*. ed. by Haklay, M. M. chapter 5, pp. 91-106, John Wiley & Sons, UK.
- Imhof, E. (1975) 'Positioning Names on Maps', **The American Cartographer**, 2(2). pp. 128-144.
- Kraak, M.J and Ormeling, F. (2010) **Cartography. Visualization of Spatial Data**, 3rd edition, Person Prentice Hall, Harlow
- Krygier, J. and Wood, D. (2011) **Making Maps, A Visual Guide to Map Design for GIS**, 2nd edition, THE Guilford Press, New York.
- Lleras, A. and Von Mühlen, A. (2003) 'Spatial Context and Top-Down Strategies in Visual Search', **Spatial Vision**, 17(4). pp. 465-482.
- MacEachren, A. M. (1982) 'The Role of Complexity and Symbolization Method in Thematic Map Effectiveness', **Annals of the Association of American Geographers**, 72(4), pp. 459-513.
- Nielsen, J. (1993) **Usability Engineering**, Academic Press, San Francisco.
- Nivala, A. M., Brewster, S. and Sarjakoski, T. (2008) 'Usability Evaluation of Web Mapping Sites' **The Cartographic Journal**. 45 (2), pp. 129-138.
- Ooms, K., Andrienko, G., Andrienko, N., De Maeyer, P. and Fack, V. (2012a) 'Analysing the Spatial Dimension of Eye Movement Data Using a Visual Analytic Approach', **Expert Systems with Applications**, 39 (11), pp. 1324–1332.
- Peterson, G. N. (2009) **GIS cartography : a guide to effective map design**, CRC Press, Taylor & Francis Group, USA.
- Phillips, R.J. (1981) 'Search for Target in a Random Arrangement of Names: an Eye Fixation Analysis', **Canadian Journal of Psychology**, 35(4), pp. 330-346.
- Rubin, J. and Chisnell, D. (2008) **Hand Book of Usability Testing: How to Plan, Design, and Conduct Effective Tests**, 2nd edition, Wiley, Indianapolis.
- van den Worm, J. (2001). 'Web Map Design in Practice', In: **Web Cartography**, ed. by Kraak, M.J. and Brown, A. chapter 7, pp. 87-107, Taylor & Francis, Great Britain.
- Wood, C.H. (2000) 'A Descriptive and Illustrated Guide For Type Placement on Small Scale Maps', **The Cartographic Journal**. 37 (1), pp. 5-18.